

Základy biostatistiky

(MD710P09)

ak. rok 2007/2008

Karel Zvára

karel.zvara@mff.cuni.cz

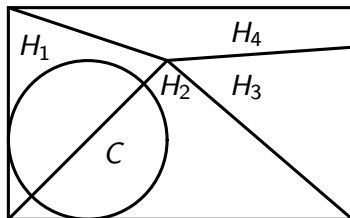
<http://www.karlin.mff.cuni.cz/~zvara>

katedra pravděpodobnosti a matematické statistiky MFF UK

(naposledy upraveno 11. března 2008)



vzorec pro úplnou pst, Bayesův vzorec

počítáme $P(H_1|C)$, např. C – správná odpověď, H_j – správná známka j 

$$P(H_1) = 0,231$$

$$P(H_2) = 0,375$$

$$P(H_3) = 0,219$$

$$P(H_4) = 0,175$$

$$P(C|H_1) = 0,589$$

$$P(C|H_2) = 0,362 \quad (\text{proč je } P(C|H_2) < P(C|H_1)?)$$

$$P(C) = P(C \cap H_1) + P(C \cap H_2)$$

$$P(C \cap H_1) = P(C|H_1)P(H_1)$$

$$P(H_1 \cap C) = P(H_1|C)P(C)$$

$$P(H_1|C) = \frac{P(H_1 \cap C)}{P(C)} = \frac{P(C|H_1)P(H_1)}{P(C|H_1)P(H_1) + P(C|H_2)P(H_2)} = \frac{1}{2}$$

obecný vzorec pro úplnou pravděpodobnost

(totéž, ale obecně)

- ▶ H_1, \dots, H_k neslučitelné (tj. $H_i \cap H_j = \emptyset$ pro $i \neq j$)
- ▶ sjednocení H_1, \dots, H_k dá jev jistý (tj. $H_1 \cup \dots \cup H_k = \Omega$)

z definice podmíněné psti plyne $P(C \cap H_j) = P(C|H_j) \cdot P(H_j)$

$$\begin{aligned} P(C) &= P(C \cap \Omega) = P(C \cap (H_1 \cup H_2 \cup \dots \cup H_k)) \\ &= P((C \cap H_1) \cup (C \cap H_2) \cup \dots \cup (C \cap H_k)) \text{ (neslučitelné jevy)} \\ &= P(C \cap H_1) + P(C \cap H_2) + \dots + P(C \cap H_k) \\ &= P(C|H_1)P(H_1) + P(C|H_2)P(H_2) + \dots + P(C|H_k)P(H_k) \end{aligned}$$

tedy obecně

$$P(C) = \sum_{j=1}^k P(C|H_j)P(H_j)$$

Bayesův vzorec

stejné předpoklady: H_j neslučitelné, sjednocení všech jistý jev [Bayes]

$$P(H_i|C) = \frac{P(H_i \cap C)}{P(C)}, \quad P(C|H_i) = \frac{P(C \cap H_i)}{P(H_i)}$$

odtud lze $P(H_i \cap C) = P(C \cap H_i)$ vyjádřit dvěma způsoby:

$$\begin{aligned} P(H_i \cap C) &= P(H_i|C)P(C) \\ &= P(C|H_i)P(H_i) \end{aligned}$$

proto pro každé i , $i = 1, \dots, k$ platí

$$P(H_i|C) = \frac{P(H_i \cap C)}{P(C)} = \frac{P(C|H_i)P(H_i)}{P(C)} = \frac{P(C|H_i)P(H_i)}{\sum_{j=1}^k P(C|H_j)P(H_j)}$$

H_1, \dots, H_k – **hypotézy**, $P(H_1|C), \dots, P(H_k|C)$ – **aposteriori** psti
 $P(H_1), \dots, P(H_k)$ – **apriorní** psti (nutně $P(H_1) + \dots + P(H_k) = 1$)

příklad: zkoušení

H_j – student si zaslouží známku j , učitel studenta (tedy j) nezná

C – student správně odpoví na položenou otázku

$P(H_j)$ – apriorní představa učitele o neznámém studentovi

$P(C|H_j)$ – obtížnost otázky, volí učitel

H_j	$P(H_j)$	$P(C H_j)$	$P(H_j)P(C H_j)$	$P(H_j C)$	$P(H_j C_2)$	$P(H_j C_3)$
1	0,20	1,00	0,2000	0,2694	0,3451	0,4230
2	0,35	0,80	0,2800	0,3771	0,3865	0,3790
3	0,25	0,65	0,1625	0,2189	0,1822	0,1452
4	0,20	0,50	0,1000	0,1347	0,0863	0,0529
Σ	1,00		0,7425	1,0000	1,0000	1,0000

$$P(C) = 0,7425$$

podobně C_2, C_3 správné odpovědi na další stejně obtížné otázky,
když použijeme předchozí aposteriorní psti jako apriorní

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psťi hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (psťní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psťi hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (psťní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psťi hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (psťní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psťi hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (psťní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psťi hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (psťní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

náhodná veličina

[random variable]

- ▶ číselně vyjádřený výsledek náhodného pokusu
- ▶ předem nevíme, který výsledek vyjde, známe jen
 - ▶ možné hodnoty
 - ▶ jejich pravděpodobnosti
- ▶ každému elementárnímu jevu přiřadíme reálné číslo
- ▶ **diskrétní rozdělení** náhodné veličiny X
 - ▶ model pro počty případů (četnosti)
 - ▶ možné hodnoty x_1^*, x_2^*, \dots
 - ▶ psti hodnot $P(X = x_1^*), P(X = x_2^*), \dots$ (pstní funkce)
- ▶ **spojité rozdělení** náhodné veličiny X
 - ▶ model pro spojitou veličiny (délka, váha, koncentrace ...)
 - ▶ obor (množina) možných hodnot X
 - ▶ hustota $f(x)$

příklad: rodina

náhodná veličina X – počet děvčatrozdělení X dáno hodnotami x_j^* a psmi těchto hodnot $P(X = x_j^*)$

ω_i	x_i	x_j^*
(m, m, m)	0	0
(m, m, f)	1	
(m, f, m)	1	1
(f, m, m)	1	
(f, f, m)	2	
(f, m, f)	2	2
(m, f, f)	2	
(f, f, f)	3	3

j	x_j^*	m_j	$P(X = x_j^*)$
1	0	1	1/8
2	1	3	3/8
3	2	3	3/8
4	3	1	1/8
součet		8	8/8

$$m = \sum_{j=1}^4 m_j = 8$$

distribuční funkce

protějšek empirické distribuční funkce (str. 15), [(cumulative) distribution function]

- ▶ pst, že X nepřekročí x

$$F_X(x) = P(X \leq x)$$

- ▶ diskrétní rozdělení:

$$F(x) = \sum_{t \leq x} P(X = t)$$

- ▶ spojité rozdělení: $F(x) = \int_{-\infty}^x f(t)dt$, kde $f(x) = \frac{dF(x)}{dx}$
- ▶ vlastnosti distribuční funkce

$$0 \leq F(x) \leq 1$$

neklesající: $x_1 < x_2 \Rightarrow F(x_2) \geq F(x_1)$

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

$$P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$

$$F(x_2) = F(x_1) + P(x_1 < X \leq x_2)$$

distribuční funkce

protějšek empirické distribuční funkce (str. 15), [(cumulative) distribution function]

- ▶ pst, že X nepřekročí x

$$F_X(x) = P(X \leq x)$$

- ▶ diskrétní rozdělení:

$$F(x) = \sum_{t \leq x} P(X = t)$$

- ▶ spojité rozdělení: $F(x) = \int_{-\infty}^x f(t)dt$, kde $f(x) = \frac{dF(x)}{dx}$
- ▶ vlastnosti distribuční funkce

$$0 \leq F(x) \leq 1$$

neklesající: $x_1 < x_2 \Rightarrow F(x_2) \geq F(x_1)$

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

$$P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$

$$F(x_2) = F(x_1) + P(x_1 < X \leq x_2)$$

distribuční funkce

protějšek empirické distribuční funkce (str. 15), [(cumulative) distribution function]

- ▶ pst, že X nepřekročí x

$$F_X(x) = P(X \leq x)$$

- ▶ diskrétní rozdělení:

$$F(x) = \sum_{t \leq x} P(X = t)$$

- ▶ spojité rozdělení: $F(x) = \int_{-\infty}^x f(t)dt$, kde $f(x) = \frac{dF(x)}{dx}$

- ▶ vlastnosti distribuční funkce

$$0 \leq F(x) \leq 1$$

neklesající: $x_1 < x_2 \Rightarrow F(x_2) \geq F(x_1)$

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

$$P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$

$$F(x_2) = F(x_1) + P(x_1 < X \leq x_2)$$

distribuční funkce

protějšek empirické distribuční funkce (str. 15), [(cumulative) distribution function]

- ▶ pst, že X nepřekročí x

$$F_X(x) = P(X \leq x)$$

- ▶ diskrétní rozdělení:

$$F(x) = \sum_{t \leq x} P(X = t)$$

- ▶ spojité rozdělení: $F(x) = \int_{-\infty}^x f(t)dt$, kde $f(x) = \frac{dF(x)}{dx}$
- ▶ vlastnosti distribuční funkce

$$0 \leq F(x) \leq 1$$

neklesající: $x_1 < x_2 \Rightarrow F(x_2) \geq F(x_1)$

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

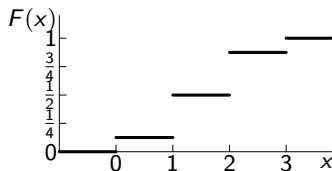
$$P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$

$$F(x_2) = F(x_1) + P(x_1 < X \leq x_2)$$

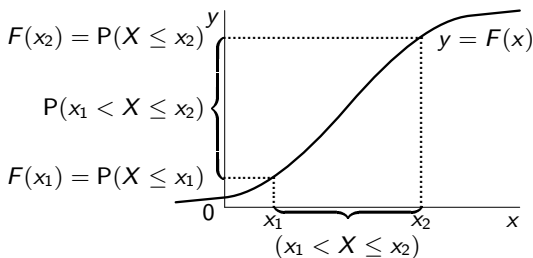
příklad diskrétního rozdělení

rozdělení počtu děvčat

j	x_j^*	$P(X = x_j^*)$	$F_X(x_j^*)$
1	0	$1/8$	$1/8$
2	1	$3/8$	$4/8$
3	2	$3/8$	$7/8$
4	3	$1/8$	$8/8$
součet		$8/8$	



geometrický význam distribuční funkce



$$P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$

$$F(x_2) = F(x_1) + P(x_1 < X \leq x_2)$$

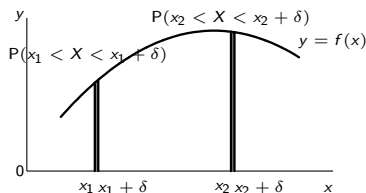
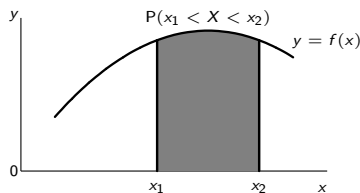
$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

hustota spojitého rozdělení

[density function]

- ▶ plocha pod celou hustotou je rovna jedné
- ▶ plocha pod hustotou nad intervalem x_1, x_2 je rovna pravděpodobnosti, že X je mezi x_1, x_2

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x) dx = 1$$

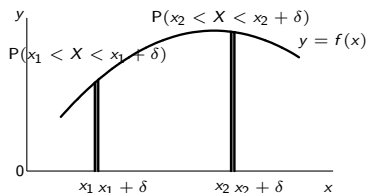
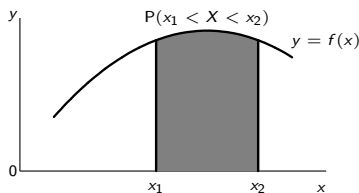


hustota spojitého rozdělení

[density function]

- ▶ plocha pod celou hustotou je rovna jedné
- ▶ plocha pod hustotou nad intervalem x_1, x_2 je rovna pravděpodobnosti, že X je mezi x_1, x_2

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x) dx = 1$$

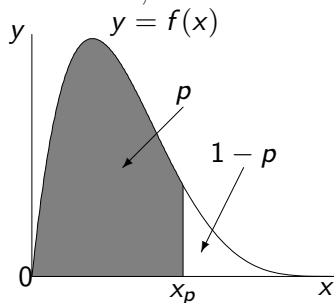
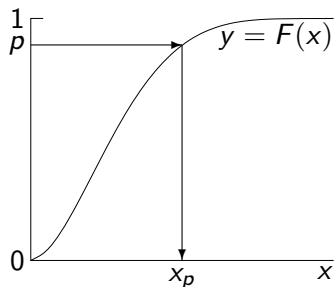


p -kvantil x_p

x_p je hodnota, pod kterou je $100p$ procent pravděpodobnosti

$$P(X \leq x_p) = p$$

např. `[qnorm(0.975)]` dá $1,959964 \doteq 1,96$

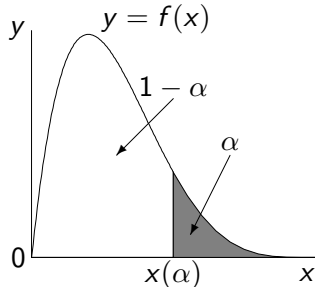
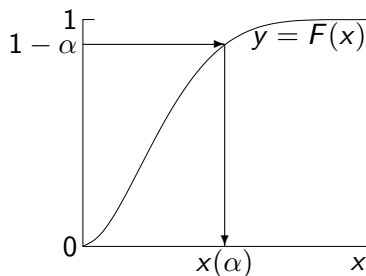


kritická hodnota $x(\alpha)$

kritická hodnota $x(\alpha)$ je překročena s pstí α

$$P(X \geq x(\alpha)) = \alpha$$

např. `[qnorm(1-0.025)]` dá $1,959964 \doteq 1,96$



střední hodnota

pokračujeme v idealizovaných představách [expected value, mean value]

- ▶ míra polohy, **populační průměr**
- ▶ metoda výpočtu se značí $E X$
- ▶ vypočtená hodnota se značí μ nebo úplněji μ_X
- ▶ **vážený průměr možných hodnot**
- ▶ ideální protějšek výběrového průměru
- ▶ diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\mu_X = E X = \sum_j x_j^* P(X = x_j^*)$$

- ▶ spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = E X = \int_{-\infty}^{\infty} x f_X(x) dx$$

střední hodnota

pokračujeme v idealizovaných představách [expected value, mean value]

- ▶ míra polohy, **populační průměr**
- ▶ metoda výpočtu se značí $E X$
- ▶ vypočtená hodnota se značí μ nebo úplněji μ_X
- ▶ **vážený průměr možných hodnot**
- ▶ ideální protějšek výběrového průměru
- ▶ diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\mu_X = E X = \sum_j x_j^* P(X = x_j^*)$$

- ▶ spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = E X = \int_{-\infty}^{\infty} x f_X(x) dx$$

střední hodnota

pokračujeme v idealizovaných představách [expected value, mean value]

- ▶ míra polohy, **populační průměr**
- ▶ metoda výpočtu se značí $E X$
- ▶ vypočtená hodnota se značí μ nebo úplněji μ_X
- ▶ **vážený průměr možných hodnot**
- ▶ ideální protějšek výběrového průměru
- ▶ diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\mu_X = E X = \sum_j x_j^* P(X = x_j^*)$$

- ▶ spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = E X = \int_{-\infty}^{\infty} x f_X(x) dx$$

střední hodnota

pokračujeme v idealizovaných představách [expected value, mean value]

- ▶ míra polohy, **populační průměr**
- ▶ metoda výpočtu se značí $E X$
- ▶ vypočtená hodnota se značí μ nebo úplněji μ_X
- ▶ **vážený průměr možných hodnot**
- ▶ ideální protějšek výběrového průměru
- ▶ diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\mu_X = E X = \sum_j x_j^* P(X = x_j^*)$$

- ▶ spojitě rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = E X = \int_{-\infty}^{\infty} x f_X(x) dx$$

střední hodnota

pokračujeme v idealizovaných představách [expected value, mean value]

- ▶ míra polohy, **populační průměr**
- ▶ metoda výpočtu se značí $E X$
- ▶ vypočtená hodnota se značí μ nebo úplněji μ_X
- ▶ **vážený průměr možných hodnot**
- ▶ ideální protějšek výběrového průměru
- ▶ diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\mu_X = E X = \sum_j x_j^* P(X = x_j^*)$$

- ▶ spojitě rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = E X = \int_{-\infty}^{\infty} x f_X(x) dx$$

střední hodnota

pokračujeme v idealizovaných představách [expected value, mean value]

- ▶ míra polohy, **populační průměr**
- ▶ metoda výpočtu se značí $E X$
- ▶ vypočtená hodnota se značí μ nebo úplněji μ_X
- ▶ **vážený průměr možných hodnot**
- ▶ ideální protějšek výběrového průměru
- ▶ diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\mu_X = E X = \sum_j x_j^* P(X = x_j^*)$$

- ▶ spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = E X = \int_{-\infty}^{\infty} x f_X(x) dx$$

střední hodnota

pokračujeme v idealizovaných představách [expected value, mean value]

- ▶ míra polohy, **populační průměr**
- ▶ metoda výpočtu se značí $E X$
- ▶ vypočtená hodnota se značí μ nebo úplněji μ_X
- ▶ **vážený průměr možných hodnot**
- ▶ ideální protějšek výběrového průměru
- ▶ diskrétní rozdělení: vahami jsou pravděpodobnosti

$$\mu_X = E X = \sum_j x_j^* P(X = x_j^*)$$

- ▶ spojité rozdělení: místo vah je hustota $f_X(x)$

$$\mu_X = E X = \int_{-\infty}^{\infty} x f_X(x) dx$$

příklad rodina

X – počet děvčat mezi třemi dětmi

j	x_j^*	$P(X = x_j^*)$	$x_j^* \cdot P(X = x_j^*)$
1	0	0,125	0,000
2	1	0,375	0,375
3	2	0,375	0,750
4	3	0,125	0,375
součet		1,000	1,500

$$\begin{aligned}
 \mu_X &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} \\
 &= 0 \cdot 0,125 + 1 \cdot 0,375 + 2 \cdot 0,375 + 3 \cdot 0,125 \\
 &= 1,5
 \end{aligned}$$

rozptyl σ^2 , směrodatná odchylka σ

[variance, standard deviation]

- ▶ míra variability, **populační rozptyl, popul. směr. odchylka**
- ▶ udává velikost kolísání (variabilitu) kolem střední hodnoty
- ▶ metoda výpočtu se značí $\text{var } X$
- ▶ vypočtená hodnota σ^2 , úplněji σ_X^2
- ▶ lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \text{var } X = E (X - \mu_X)^2 = E (X^2) - (\mu_X)^2$$

- ▶ ideální protějšky výběrového rozptylu, směr. odchylky
- ▶ **diskrétní rozdělení**

$$\sigma_X^2 = \text{var } X = \sum_j (x_j^* - \mu_X)^2 P (X = x_j^*)$$

- ▶ **spojité rozdělení** $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$

rozptyl σ^2 , směrodatná odchylka σ

[variance, standard deviation]

- ▶ míra variability, **populační rozptyl, popul. směr. odchylka**
- ▶ udává velikost kolísání (variabilitu) kolem střední hodnoty
- ▶ metoda výpočtu se značí $\text{var } X$
- ▶ vypočtená hodnota σ^2 , úplněji σ_X^2
- ▶ lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \text{var } X = E (X - \mu_X)^2 = E (X^2) - (\mu_X)^2$$

- ▶ ideální protějšky výběrového rozptylu, směr. odchylky
- ▶ **diskrétní rozdělení**

$$\sigma_X^2 = \text{var } X = \sum_j (x_j^* - \mu_X)^2 P (X = x_j^*)$$

- ▶ **spojité rozdělení** $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$

rozptyl σ^2 , směrodatná odchylka σ

[variance, standard deviation]

- ▶ míra variability, **populační rozptyl, popul. směr. odchylka**
- ▶ udává velikost kolísání (variabilitu) kolem střední hodnoty
- ▶ metoda výpočtu se značí $\text{var } X$
- ▶ vypočtená hodnota σ^2 , úplněji σ_X^2
- ▶ lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \text{var } X = E (X - \mu_X)^2 = E (X^2) - (\mu_X)^2$$

- ▶ ideální protějšky výběrového rozptylu, směr. odchylky
- ▶ **diskrétní rozdělení**

$$\sigma_X^2 = \text{var } X = \sum_j (x_j^* - \mu_X)^2 P (X = x_j^*)$$

- ▶ **spojité rozdělení** $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$

rozptyl σ^2 , směrodatná odchylka σ

[variance, standard deviation]

- ▶ míra variability, **populační rozptyl, popul. směr. odchylka**
- ▶ udává velikost kolísání (variabilitu) kolem střední hodnoty
- ▶ metoda výpočtu se značí $\text{var } X$
- ▶ vypočtená hodnota σ^2 , úplněji σ_X^2
- ▶ lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \text{var } X = E (X - \mu_X)^2 = E (X^2) - (\mu_X)^2$$

- ▶ ideální protějšky výběrového rozptylu, směr. odchylky
- ▶ **diskrétní rozdělení**

$$\sigma_X^2 = \text{var } X = \sum_j (x_j^* - \mu_X)^2 P (X = x_j^*)$$

- ▶ **spojité rozdělení** $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$

rozptyl σ^2 , směrodatná odchylka σ

[variance, standard deviation]

- ▶ míra variability, **populační rozptyl, popul. směr. odchylka**
- ▶ udává velikost kolísání (variabilitu) kolem střední hodnoty
- ▶ metoda výpočtu se značí $\text{var } X$
- ▶ vypočtená hodnota σ^2 , úplněji σ_X^2
- ▶ lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \text{var } X = E (X - \mu_X)^2 = E (X^2) - (\mu_X)^2$$

- ▶ ideální protějšky výběrového rozptylu, směr. odchylky
- ▶ **diskrétní rozdělení**

$$\sigma_X^2 = \text{var } X = \sum_j (x_j^* - \mu_X)^2 P (X = x_j^*)$$

- ▶ **spojité rozdělení** $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$

rozptyl σ^2 , směrodatná odchylka σ

[variance, standard deviation]

- ▶ míra variability, **populační rozptyl, popul. směr. odchylka**
- ▶ udává velikost kolísání (variabilitu) kolem střední hodnoty
- ▶ metoda výpočtu se značí $\text{var } X$
- ▶ vypočtená hodnota σ^2 , úplněji σ_X^2
- ▶ lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \text{var } X = E (X - \mu_X)^2 = E (X^2) - (\mu_X)^2$$

- ▶ ideální protějšky výběrového rozptylu, směr. odchylky
- ▶ **diskrétní rozdělení**

$$\sigma_X^2 = \text{var } X = \sum_j (x_j^* - \mu_X)^2 P (X = x_j^*)$$

- ▶ **spojité rozdělení** $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$

rozptyl σ^2 , směrodatná odchylka σ

[variance, standard deviation]

- ▶ míra variability, **populační rozptyl, popul. směr. odchylka**
- ▶ udává velikost kolísání (variabilitu) kolem střední hodnoty
- ▶ metoda výpočtu se značí $\text{var } X$
- ▶ vypočtená hodnota σ^2 , úplněji σ_X^2
- ▶ lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \text{var } X = E (X - \mu_X)^2 = E (X^2) - (\mu_X)^2$$

- ▶ ideální protějšky výběrového rozptylu, směr. odchylky
- ▶ **diskrétní rozdělení**

$$\sigma_X^2 = \text{var } X = \sum_j (x_j^* - \mu_X)^2 P (X = x_j^*)$$

- ▶ **spojité rozdělení** $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$

rozptyl σ^2 , směrodatná odchylka σ

[variance, standard deviation]

- ▶ míra variability, **populační rozptyl, popul. směr. odchylka**
- ▶ udává velikost kolísání (variabilitu) kolem střední hodnoty
- ▶ metoda výpočtu se značí $\text{var } X$
- ▶ vypočtená hodnota σ^2 , úplněji σ_X^2
- ▶ lze vyjádřit pomocí střední hodnoty

$$\sigma_X^2 = \text{var } X = E (X - \mu_X)^2 = E (X^2) - (\mu_X)^2$$

- ▶ ideální protějšky výběrového rozptylu, směr. odchylky
- ▶ **diskrétní rozdělení**

$$\sigma_X^2 = \text{var } X = \sum_j (x_j^* - \mu_X)^2 P (X = x_j^*)$$

- ▶ **spojité rozdělení** $\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$

příklad rodina

X – počet děvčat mezi třemi dětmi, $\mu_X = 1,5$

j	x_j^*	p_j	$x_j^* - \mu_X$	$(x_j^* - \mu_X)^2$	$(x_j^* - \mu_X)^2 p_j$
1	0	0,125	-1,5	2,25	0,28125
2	1	0,375	-0,5	0,25	0,09375
3	2	0,375	0,5	0,25	0,09375
4	3	0,125	1,5	2,25	0,28125
Σ		1,000	0,0		0,75000

$$\begin{aligned}\sigma_X^2 &= \sum_j (x_j^* - \mu_X)^2 p_j \\ &= (0 - 1,5)^2 \cdot 0,125 + (1 - 1,5)^2 \cdot 0,375 \\ &\quad + (2 - 1,5)^2 \cdot 0,375 + (3 - 1,5)^2 \cdot 0,125 = 0,75 \\ \sigma_X &= \sqrt{0,75} = 0,866025\end{aligned}$$

sdružené rozdělení

- ▶ abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice, ...) náhodných veličin, tedy chování **náhodného vektoru**
- ▶ příklad **rodina**
 - ▶ X – počet děvčat v rodině s třemi dětmi
 - ▶ Y – počet děvčat mezi dvěma staršími dětmi
 - ▶ Z – počet hochů v rodině s třemi dětmi
- ▶ zajímá nás rozdělení náhodného vektoru (X, Y)
- ▶ proč nemá smysl vyšetřovat **vektor** (X, Z) ?
- ▶ (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- ▶ abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice, ...) náhodných veličin, tedy chování **náhodného vektoru**
- ▶ příklad **rodina**
 - ▶ X – počet děvčat v rodině s třemi dětmi
 - ▶ Y – počet děvčat mezi dvěma staršími dětmi
 - ▶ Z – počet hochů v rodině s třemi dětmi
- ▶ zajímá nás rozdělení náhodného vektoru (X, Y)
- ▶ proč nemá smysl vyšetřovat **vektor** (X, Z) ?
- ▶ (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- ▶ abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice, ...) náhodných veličin, tedy chování **náhodného vektoru**
- ▶ příklad **rodina**
 - ▶ X – počet děvčat v rodině s třemi dětmi
 - ▶ Y – počet děvčat mezi dvěma staršími dětmi
 - ▶ Z – počet hochů v rodině s třemi dětmi
- ▶ zajímá nás rozdělení náhodného vektoru (X, Y)
- ▶ proč nemá smysl vyšetřovat **vektor** (X, Z) ?
- ▶ (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- ▶ abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice, ...) náhodných veličin, tedy chování **náhodného vektoru**
- ▶ příklad **rodina**
 - ▶ X – počet děvčat v rodině s třemi dětmi
 - ▶ Y – počet děvčat mezi dvěma staršími dětmi
 - ▶ Z – počet hochů v rodině s třemi dětmi
- ▶ zajímá nás rozdělení náhodného vektoru (X, Y)
- ▶ proč nemá smysl vyšetřovat **vektor** (X, Z) ?
- ▶ (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- ▶ abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice, ...) náhodných veličin, tedy chování **náhodného vektoru**
- ▶ příklad **rodina**
 - ▶ X – počet děvčat v rodině s třemi dětmi
 - ▶ Y – počet děvčat mezi dvěma staršími dětmi
 - ▶ Z – počet hochů v rodině s třemi dětmi
- ▶ zajímá nás rozdělení náhodného vektoru (X, Y)
- ▶ proč nemá smysl vyšetřovat **vektor** (X, Z) ?
- ▶ (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- ▶ abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice, ...) náhodných veličin, tedy chování **náhodného vektoru**
- ▶ příklad **rodina**
 - ▶ X – počet děvčat v rodině s třemi dětmi
 - ▶ Y – počet děvčat mezi dvěma staršími dětmi
 - ▶ Z – počet hochů v rodině s třemi dětmi
- ▶ zajímá nás rozdělení náhodného vektoru (X, Y)
- ▶ proč nemá smysl vyšetřovat **vektor** (X, Z) ?
- ▶ (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- ▶ abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice, ...) náhodných veličin, tedy chování **náhodného vektoru**
- ▶ příklad **rodina**
 - ▶ X – počet děvčat v rodině s třemi dětmi
 - ▶ Y – počet děvčat mezi dvěma staršími dětmi
 - ▶ Z – počet hochů v rodině s třemi dětmi
- ▶ zajímá nás rozdělení náhodného vektoru (X, Y)
- ▶ proč nemá smysl vyšetřovat **vektor** (X, Z) ?
- ▶ (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené rozdělení

- ▶ abychom mohli popsat **závislost** náhodných veličin, zajímáme se o **společné** chování dvojice (trojice, ...) náhodných veličin, tedy chování **náhodného vektoru**
- ▶ příklad **rodina**
 - ▶ X – počet děvčat v rodině s třemi dětmi
 - ▶ Y – počet děvčat mezi dvěma staršími dětmi
 - ▶ Z – počet hochů v rodině s třemi dětmi
- ▶ zajímá nás rozdělení náhodného vektoru (X, Y)
- ▶ proč nemá smysl vyšetřovat **vektor** (X, Z) ?
- ▶ (protože Z je určeno X jednoznačně: $Z = 3 - X$)

sdružené, marginální a podmíněné rozdělení

sdružené rozdělení – popisuje **společné chování** X, Y

$$P(X = x_i^*, Y = y_j^*) \text{ resp. } f_{X,Y}(x, y)$$

marginální rozdělení: chování jedné bez ohledu na hodnotu druhé

$$P(X = x_i^*) = \sum_j P(X = x_i^*, Y = y_j^*) \quad \forall x_i^*$$

$$P(Y = y_j^*) = \sum_i P(X = x_i^*, Y = y_j^*) \quad \forall y_j^*$$

podmíněné rozdělení: chování X při **dané** hodnotě Y

$$P(Y = y_j^* | X = x_i^*) = \frac{P(X = x_i^*, Y = y_j^*)}{P(X = x_i^*)}$$

příklad rodina

X počet děvčat, Y počet hochů mezi třemi dětmi

sdružené, marginální a podmíněné rozdělení

ω_i	x_i	y_i
(m, m, m)	0	0
(m, m, f)	1	1
(m, f, m)	1	1
(f, m, m)	1	0
(f, f, m)	2	1
(f, m, f)	2	1
(m, f, f)	2	2
(f, f, f)	3	2

x_i^*	y_j^*			celkem
	0	1	2	
0	1/8	0	0	1/8
1	1/8	2/8	0	3/8
2	0	2/8	1/8	3/8
3	0	0	1/8	1/8
	2/8	4/8	2/8	1

x_i^*	y_j^*			celkem
	0	1	2	
0	1	0	0	1
1	1/3	2/3	0	1
2	0	2/3	1/3	1
3	0	0	1	1

kovariance

protějšek s_{xy} , [covariance]

kovariance vyjadřuje vzájemnou závislost náhodných veličin:

$$\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y)$$

$$\sigma_{X,Y} = \sum_i \sum_j (x_i^* - \mu_X)(y_j^* - \mu_Y)P(X = x_i^*, Y = y_j^*)$$

označení metody výpočtu: $\text{cov}(X, Y)$

zřejmě platí $\text{cov}(X, X) = \text{var } X$ tj. $\sigma_{X,X} = \sigma_X^2$

pro **nezávislé** náhodné veličiny platí

(ze znalosti hodnoty jedné nic nevíme o druhé)

$$P(X = x_i^*, Y = y_j^*) = P(X = x_i^*) \cdot P(Y = y_j^*), \quad \forall(x_i^*, y_j^*)$$

jsou-li X, Y – **nezávislé** $\Rightarrow \sigma_{X,Y} = 0$ (nikoliv obrácená implikace)

příklad rodina

x_i^*	y_j^*			celkem
	0	1	2	
0	0,125	0	0	0,125
1	0,125	0,250	0	0,375
2	0	0,250	0,125	0,375
3	0	0	0,125	0,125
celkem	0,250	0,500	0,250	1,000

$$\mu_X = 0 \cdot 0,125 + 1 \cdot 0,375 + 2 \cdot 0,375 + 3 \cdot 0,125 = 1,5$$

$$\mu_Y = 0 \cdot 0,250 + 1 \cdot 0,500 + 2 \cdot 0,250 = 1$$

$$\sigma_X^2 = (0 - 1,5)^2 \cdot 0,125 + \dots + (3 - 1,5)^2 \cdot 0,125 = 0,75$$

$$\sigma_Y^2 = (0 - 1)^2 \cdot 0,25 + (1 - 1)^2 \cdot 0,5 + (2 - 1)^2 \cdot 0,25 = 0,5$$

$$\sigma_{XY} = (0 - 1,5) \cdot (0 - 1) \cdot 0,125 + \dots = 0,5$$

X, Y jsou závislé, neboť např. $0,25 \cdot 0,125 \neq 0,125$